OJPHI

ISDS 2013 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

From Noise to Characterization Tool: Assessing Biases in Influenza Surveillance Methods Using a Bayesian Hierarchical Model

Ying Zhang*[1], Ali Arab[1], Michael A. Stoto[1] and Bejamin J. Cowling[2]

[1]Georgetown University, Reston, VA, USA; [2]The University of Hong Kong, Hong Kong, Hong Kong

Objective

Our goal is to develop a statistical framework to characterize influenza surveillance systems and their sensitivity to information environment.

Introduction

Infectious disease surveillance is a process, the product of which reflects both real illness and public awareness of the disease (Figure 1). According to our previous research studies [1,2], decisions made by patients, healthcare providers, and public health professionals about seeking and providing healthcare and about reporting cases to health authorities are all influenced by the information environment, which changes constantly. Biases are therefore imbedded in each surveillance systems, and need to be assessed to provide better situational awareness for decision-making.

Methods

We identified influenza surveillance data from Hong Kong covering health care providers, laboratories and residential care homes for the elderly. A Bayesian hierarchical model was developed to estimate the statistical relationships between influenza surveillance data and information environment data (e.g. HealthMap, Google).

For data in percentages:

<u>Data model</u>: [*data|process,data parameters*]

$Log(Y_{j,t})\sim N(\mu_{j,t},\sigma_j^2)$

<u>Process model</u>: [*process|process parameters*]

$\mu_{j,t}=\theta_{j,t}*X_t+\phi_{j,t}$

$\mu_{j,t}=(\beta_{j,t,1}+\beta_{j,t,2}*k^p{}_{1,t}+\ldots+\beta_{j,t,m}*k^p{}_{m-1,t})*X_t+\alpha_{j,t,1}+\alpha_{j,t,2}*k^p{}_{m,t}+\ldots+\alpha_{j,t,n}*k^p{}_{n-1,t}+\varepsilon_t$

$\varepsilon_t\sim N(0,\sigma_j^2)$

For data counts:

<u>Data model</u>: [*data|process,data parameters*]

$Y_{j,t}\sim Pois(\lambda_{j,t})$

<u>Process model</u>: [*process|process parameters*]

$Log(\lambda_{j,t})=\theta_{j,t}*X_t+\phi_{j,t}$

$Log(\lambda_{j,t})=(\beta_{j,t,1}+\beta_{j,t,2}*k^p{}_{1,t}+\ldots+\beta_{j,t,m}*k^p{}_{m-1,t})*X_t+\alpha_{j,t,1}+\alpha_{j,t,2}*k^p{}_{m,t}+\ldots+\alpha_{j,t,n}*k^p{}_{n-1,t}$

For both count and percentage data, the parameter model is:

<u>Parameter Model</u>: [*data and process parameters*]

$\beta_{j,t,m}\sim dnorm(0,.01)$

$\alpha_{j,t,n}\sim dnorm(0,.01)$

$\sigma_j^2\sim dgamma(.01,.01)$

X_t is the estimated influenza incidence rate in the whole population; $Y_{j,t}$ refers to surveillance system *j* at time *t*. $\theta_{j,t}$ describes the component of true infections that surveillance system j captures at time t, which is further fitted into a linear regression model with predictive variables of information environment index.$\phi_{j,t}$ - also fitted into a linear regression model with another set of predictors - is defined as perception bias, which estimates the proportion in the surveillance data which cannot be fully explained by the true infections. $\beta_{j,t,m}$ and $\alpha_{j,t,n}$ are the coefficients for a set of information environment $k^p{}_{l,t}$ *(l=1,...,m-1,m,...n-1)* during the pandemic period.

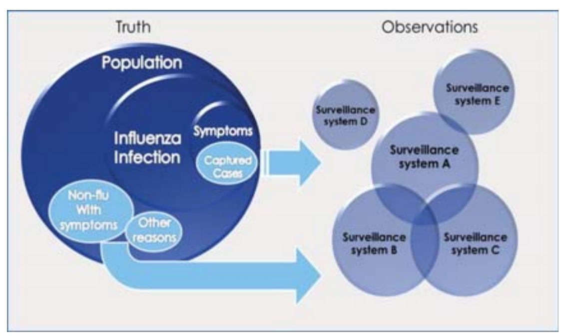
Using Markov Chain Monte-Carlo (MCMC) method in OpenBUGS, a posterior distribution was generated for every parameter to characterize each data streams.

Results

The model identified surveillance systems characteristics - percentages, broad case definition, and senior population - that are more resistant to the information environment. General practitioner (%ILI-visit) and Laboratory (%positive) systems seem to capture the true infection at a constant proportion, and are less influenced by information environment. Surveillance systems with influenza-specific case definitions tend to reflect biases of both healthcare seekers and providers.

Conclusions

The study identified the characteristics that are likely to be associated with better performance when information environment is changing rapidly. Moreover, the characterization tool, can help practitioners make a more informed decision on which surveillance systems to monitor, given their primary concerns of real illness versus public awareness.



Conceptual model for biases in influenza surveillance data

Keywords

influenza surveillance; bayesian; informational environment

Acknowledgments

This research was conducted with funding support awarded to the Harvard School of Public Health under cooperative agreements with the US CDC grant number 5P01TP000307-01. The data was provided by Centre for Health Protection and Hospital Authority in Hong Kong, and HealthMap.

References

1. Zhang Y, Lopez-Gatell H, Alpuche-Aranda CM, Stoto MA: Did Advances in Global Surveillance and Notification Systems Make a Difference in the 2009 H1N1 Pandemic?-A Retrospective Analysis. PLoS One 2013, 8(4):e59893.
2. Stoto MA: The effectiveness of U.S. public health surveillance systems for situational awareness during the 2009 H1N1 pandemic: a retrospective analysis. PLoS One 2012, 7(8):e40984.

*Ying Zhang
E-mail: yz62@georgetown.edu